

DG-PIC: Domain Generalized Point-In-Context Learning for Point Cloud Understanding

Supplementary Materials

Jincen Jiang^{1*}, Qianyu Zhou^{2*}, Yuhang Li^{1,3}, Xuequan Lu⁴,
Meili Wang⁵, Lizhuang Ma², Jian Chang¹, and Jian Jun Zhang¹

¹ National Centre for Computer Animation, Bournemouth University, Dorset, UK

² Shanghai Jiao Tong University, Shanghai, China

³ Shanghai University, Shanghai, China

⁴ La Trobe University, Victoria, Australia

⁵ Northwest A&F University, Yangling, China

{jiangj, jchang, jzhang}@bournemouth.ac.uk
zhouqianyu@sjtu.edu.cn, yuhangli@shu.edu.cn, b.lu@latrobe.edu.au
ma-lz@cs.sjtu.edu.cn, wml@nwsuaf.edu.cn

This supplementary includes the following items:

- Section A demonstrates our DG-PIC’s generalization ability under the leave-one-domain-out setting.
- Section B elaborates on the details and analysis of the hyperparameters on our DG-PIC. We investigate the impacts of different degrees of the prompt selection factor λ and mask ratio γ to verify the effectiveness of our method.
- Section C shows more ablation studies of our method, including the pooling strategies to obtain global features and prototypes.
- Section D visualizes the qualitative results of our DG-PIC, including t-SNE visualization of features from different domains, the differences (*i.e.*, gap) across domains, and results comparisons with other methods.

A Cross Verification

Following some previous DG methods [4, 10], we adopt the leave-one-domain-out setting to assess the effectiveness of our method. We designate one dataset from our benchmark as the target domain and use the remaining datasets as the source domain. The results for cross-validation are presented in Table A1, where we compare our method with PIC [1]. We can see that our method consistently achieves significantly better results in all cases, showcasing the superiority of our method. This observation confirms the effectiveness of our method, which has a strong generalization ability between different domains.

* Equal contributions.

 Corresponding authors.

Table A1: Cross verification with leave-one-domain-out protocol. The Chamfer Distance ($\times 10^{-3}$) serves as the metric for three different tasks.

Methods	Target Domain	Reconstruction	Denoising	Registration
PIC	ScanObjectNN	72.9	80.0	12.7
DG-PIC		4.1	15.2	5.8
PIC	ScanNet	21.7	49.9	61.0
DG-PIC		7.3	23.1	35.6
PIC	ModelNet	22.6	32.9	10.3
DG-PIC		7.8	19.0	6.0
PIC	ShapeNet	37.5	39.8	11.8
DG-PIC		8.5	19.3	10.0

B Hyperparameter Analysis

B.1 Prompt Selection Factor λ

As indicated in Eq. (8) in the manuscript, λ serves as the factor governing the weights assigned to global-level and local-level features when calculating feature distances. The optimal results, as observed in Table B2 (Upper), are achieved when λ is set to 0.5. This balance in weights enables the model to consider the dual-level features equally, encompassing both global shape information and local geometric structure, which facilitates the model’s effectiveness.

B.2 Mask Ratio γ

In addition to employing a high mask ratio in the Masked Point Modeling (MPM) framework [8] in the manuscript, we also discuss the effect of different mask ratios on the results. As shown in Table B2 (Bottom), lower mask ratios during pre-training diminish the model’s performance on different tasks. In contrast, a higher mask ratio, such as 0.9, limits visible data excessively, hindering effective representation. Maintaining an appropriate high mask ratio, *i.e.*, 0.7, similar to MAE [3], enables the model to learn hidden features better.

C More Ablation studies

We select max-pooling as it is a typical method for getting global features [5]. Table C3 shows *max or average pooling* for global prototypes leads to similar results. Local prototype of a patch $m \in [1, M]$ is $\frac{1}{N} \sum_{n=1}^N f_n^m$, where N is the sample (point cloud) number in a dataset (domain) and $f_n^m \in \mathbb{R}^{C \times 1}$ is patch m ’s local feature. Global prototype of a point cloud is $(\frac{1}{N} \sum_{n=1}^N \max(F_n)) \in \mathbb{R}^{C \times 1}$ where $F_n = \{f_n^m\} \in \mathbb{R}^{C \times M}$. Table C4 shows average (above $\sum_{n=1}^N$) leads to better results, as *maximize* ignores other samples in the dataset.

Table B2: Ablation study results for different hyperparameters. We use the ScanObjectNN dataset as the target domain and the other three as the source domain. We report the Chamfer Distance ($\times 10^{-3}$) as the metric for three different tasks.

Models	Hyperparameter	Reconstruction	Denoising	Registration
Model A	$\lambda = 0.1$	5.2	27.2	6.5
Model B	$\lambda = 0.3$	4.8	19.3	6.2
Our Choice	$\lambda = 0.5$	4.1	15.2	5.8
Model C	$\lambda = 0.7$	5.1	23.8	6.7
Model D	$\lambda = 0.9$	5.3	29.4	6.4
Model E	$\gamma = 0.3$	12.7	34.9	22.4
Model F	$\gamma = 0.5$	8.5	25.2	13.8
Our Choice	$\gamma = 0.7$	4.1	15.2	5.8
Model G	$\gamma = 0.9$	10.6	23.9	11.4

Table C3: Ablation on global features.

Table C4: Prototypes estimation.

Pooling	Rec./Den./Reg.
Avg pool.	4.5/ 15.0 /6.3
Max pool.	4.1 /15.2/ 5.8

Prototypes	Rec./Den./Reg.
Maximize	5.7/19.4/6.5
Average	4.1 / 15.2 / 5.8

D Visualization Analysis

D.1 More Visual Comparison

We conduct a comparison with various point cloud learning networks across three distinct tasks, *i.e.*, reconstruction, denoising, and registration, as depicted in Figure D1. The results reveal that our method produces visual results closely resembling the original ground truth. Notably, several common networks, including PointNet [5], DGCNN [6] and PCT [2], even for PointCutMix [9], which incorporates domain generalization, struggle to perform well in the multi-domain multi-task setting. Additionally, PIC [1], which also utilizes ICL, faces challenges in the multi-domain setting, leading to suboptimal results. Our method addresses multi-domain multi-task learning simultaneously, proving effective not only in overcoming the domain gap through test-time target feature shifting but also in generalizing multi-tasks through ICL, thereby achieving superior results.

D.2 Visualization of Different Domains

Figure D2 displays chairs from various domains to demonstrate the domain shifts across different sources, *i.e.*, well-structured synthetic data like ModelNet40 [7] *v.s.* noisy and incomplete real-world data such as ScanObjectNN.

D.3 T-SNE Visualization of Shifting

We visualize the learned features using t-SNE in Figure D3 to validate the effectiveness of the Test-time target feature shifting. As evident from the comparison

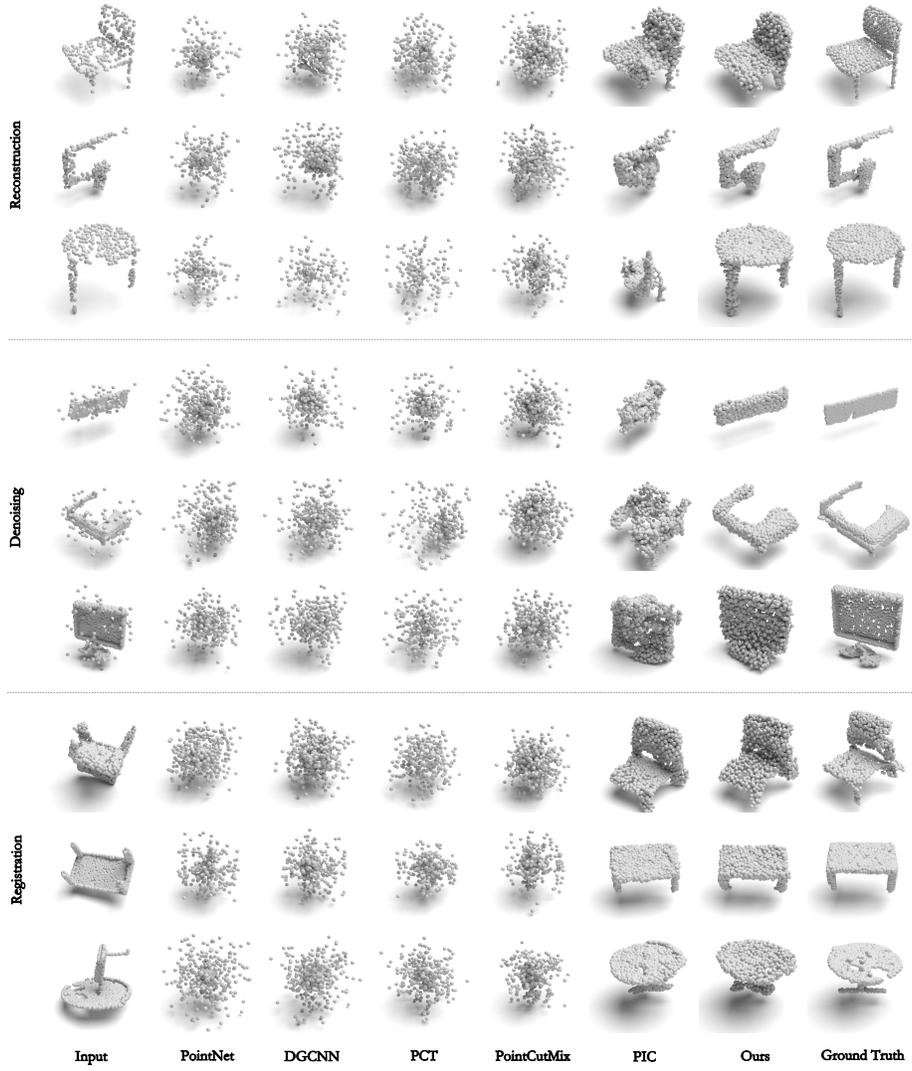


Fig. D1: Comparison of three different point cloud understanding tasks between our DG-PIC and other methods.

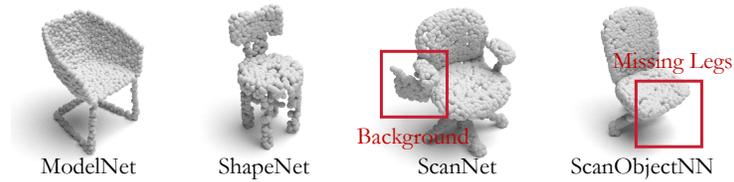


Fig. D2: Visualization of chairs from different domains.

with the case without shifting, the shifting operation successfully pulls the features of the target samples closer to the source domains that share similarities with them (illustrated by the red dots scattered within different source domains). This operation provides the model with more favorable known information, resulting in a more effective outcome.

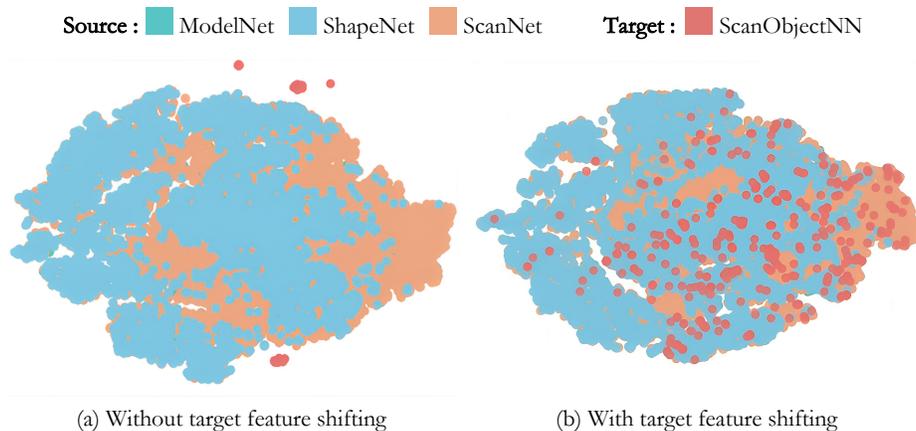


Fig. D3: T-SNE visualization of features learned by our method. The red dots (indicate the target samples) are pulled towards various source domains.

References

1. Fang, Z., Li, X., Li, X., Buhmann, J.M., Loy, C.C., Liu, M.: Explore in-context learning for 3d point cloud understanding. arXiv preprint arXiv:2306.08659 (2023)
2. Guo, M.H., Cai, J.X., Liu, Z.N., Mu, T.J., Martin, R.R., Hu, S.M.: Pct: Point cloud transformer. *Computational Visual Media* **7**, 187–199 (2021)
3. He, K., Chen, X., Xie, S., Li, Y., Dollár, P., Girshick, R.: Masked autoencoders are scalable vision learners. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. pp. 16000–16009 (2022)
4. Park, J., Han, D.J., Kim, S., Moon, J.: Test-time style shifting: Handling arbitrary styles in domain generalization. In: *International Conference on Machine Learning* (2023)
5. Qi, C.R., Su, H., Mo, K., Guibas, L.J.: Pointnet: Deep learning on point sets for 3d classification and segmentation. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. pp. 652–660 (2017)
6. Wang, Y., Sun, Y., Liu, Z., Sarma, S.E., Bronstein, M.M., Solomon, J.M.: Dynamic graph cnn for learning on point clouds. *ACM Transactions on Graphics* **38**(5), 1–12 (2019)
7. Wu, Z., Song, S., Khosla, A., Yu, F., Zhang, L., Tang, X., Xiao, J.: 3d shapenets: A deep representation for volumetric shapes. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. pp. 1912–1920 (2015)

8. Yu, X., Tang, L., Rao, Y., Huang, T., Zhou, J., Lu, J.: Point-bert: Pre-training 3d point cloud transformers with masked point modeling. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 19313–19322 (2022)
9. Zhang, J., Chen, L., Ouyang, B., Liu, B., Zhu, J., Chen, Y., Meng, Y., Wu, D.: Pointcutmix: Regularization strategy for point cloud classification. *Neurocomputing* **505**, 58–67 (2022)
10. Zhao, X., Liu, C., Sicilia, A., Hwang, S.J., Fu, Y.: Test-time fourier style calibration for domain generalization. In: The International Joint Conference on Artificial Intelligence (2022)